

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 11:11:12 ; Search time 16.02 Seconds

(without alignments)
30.238 Million cell updates/sec

Title: US-09-288-719-3

Sequence: 1 GGGGSGGRSGGGS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	86.4	465	FXD3_MOUSE	O61060 mus musculus
2	66	81.5	779	M130_STRPU	P08472 strongyloce
3	65	80.2	104	HOL3_HOUDI	Q25055 holotrichia
4	65	80.2	643	K2C1_HUMAN	P04264 homo sapien
5	63	77.8	206	TWST_MOUSE	P26687 mus musculus
6	63	77.8	214	GRP2_NICSY	P27484 nicotiana s
7	63	77.8	265	YPG1_ZYMO	P30794 zymonomas m
8	63	77.8	431	HXB3_HUMAN	P14651 homo sapien
9	63	77.8	433	HXB3_MOUSE	P09026 mus musculus
10	63	77.8	1434	KDGE_DROME	Q09103 drosophila
11	62	76.5	157	GRPA_MAIZE	P10979 zea mays (m
12	62	76.5	157	GRP_DAUCA	Q03878 daucus caro
13	62	76.5	165	GRP1_ORYSA	P25074 oryza sativ
14	62	76.5	168	GRP2_SORBI	Q09070 sorghum dic
15	62	76.5	280	CHIA_MAIZE	P29022 zea mays (m
16	62	76.5	321	PUR_MOUSE	P23669 mus musculu
17	62	76.5	322	PUR_HUMAN	Q00577 homo sapien
18	62	76.5	323	HXB_MOUSE	P23813 mus musculu
19	62	76.5	440	FXGA_CHICK	Q08937 gallus galli
20	62	76.5	495	BRN1_MOUSE	P31361 mus musculu
21	62	76.5	497	BRN1_RAT	Q61262 ratius norv
22	62	76.5	500	BRN1_HUMAN	P20264 homo sapien
23	62	76.5	539	DOP2_DROME	Q24563 drosophila
24	62	76.5	569	K1CJ_MOUSE	P02535 mus musculu
25	62	76.5	622	SR68_CANFA	Q00004 canis famli
26	62	76.5	663	DUS8_MOUSE	Q09112 mus musculu
27	62	76.5	688	BOMD_MOUSE	Q54839 mus musculu
28	62	76.5	979	REF1_HUMAN	P22670 homo sapien
29	61	76.5	1627	TP2B_CHICK	Q42131 gallus galli
30	61	75.3	316	LORI_HUMAN	P23490 homo sapien
31	60	74.1	266	CANS_RABIT	P06813 oryctolagus
32	60	74.1	332	STX3_HUMAN	Q95343 homo sapien
33	60	74.1	401	CSP_PLACC	P08674 plasmodium

34	60	74.1	410	1	BR32_HUMAN	Q12837 homo sapien
35	60	74.1	653	1	YKR2_CAEEL	P34308 caenorhabdi
36	60	74.1	722	1	COAT_PAVTH	P03136 hamster par
37	59	72.8	43	1	GRW1_LYCES	Q01157 lycopersico
38	59	72.8	151	1	GDF7_MOUSE	P43029 mus musculu
39	59	72.8	263	1	CANS_BOVIN	P13135 bos taurus
40	59	72.8	266	1	CANS_PIG	P04574 sus scrofa
41	59	72.8	268	1	CANS_HUMAN	P04632 homo sapien
42	59	72.8	268	1	EP34_HCVYA	P16768 human cytom
43	59	72.8	367	1	BET3_MESAU	Q09029 mesocricetu
44	59	72.8	387	1	SOX1_HUMAN	Q00570 homo sapien
45	59	72.8	391	1	SOX1_MOUSE	P53783 mus musculu

ALIGNMENTS

RESULT ID	FXD3_MOUSE	STANDARD:	PRT:	465 AA.
AC	061060:	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	FORKHEAD BOX PROTEIN D3 (HNF3/PH TRANSCRIPTION FACTOR GENESIS)			
DE	(HEPATOCYTE NUCLEAR FACTOR 3 FORKHEAD HOMOLOG 2) (HFH-2).			
GN	FOXD3 OR HFH2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RA	Hromas R.A., Costa R.H., Xu D., Sutton J.L.;			
RL	Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.			
RM	[2]			
RP	SEQUENCE FROM N.A.			
RA	Labosky P.A., Kaestner K.H.;			
RT	"The winged helix transcription factor Hfh2 is expressed in neural crest and spinal cord during mouse development."			
RL	Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.			
CC	- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.			
CC	- SUBCELLULAR LOCATION: NUCLEAR.			
CC	- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: U41047; AA87569.1; -			
DR	EMBL: AF067421; AAC28352.1; -			
DR	MGD: MGI:1347473; FOXD3.			
DR	INTERPRO: IPR001766; -			
DR	PFAM: PF00250; Fork_head. 1.			
DR	PRINTS: PR00053; FORKHEAD.			
DR	PROSITE: PS00657; FORK_HEAD_1; 1.			
DR	PROSITE: PS00658; FORK_HEAD_2; 1.			
DR	PROSITE: PS50039; FORK_HEAD_3; 1.			
KW	DNA-binding; Nuclear protein; Transcription regulation.			
FT	DOMAIN 106 113			
FT	DNA_BIND 131 225			
FT	DOMAIN 252 257			
FT	DOMAIN 265 270			
FT	DOMAIN 275 281			
FT	DOMAIN 380 393			
FT	DOMAIN 385 399			
FT	DOMAIN 447 457			
FT	DOMAIN 457 457			
SO	SEQUENCE 465 AA; 47092 MW; 6F8B5B3D8C7564D CRC64;			

Query Match 86.4%; Score 70; DB 1; Length 465;
 Best Local Similarity 86.7%; Pred. No. 0.52;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGGS 15
 ||||| |||||
 DB 380 GGGGSGGAGGGGGS 394

RESULT 2

M130_STRPU STANDARD; PRT: 779 AA.

ID M130_STRPU STANDARD; PRT: 779 AA.
 AC P08472; 1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 17, Last sequence update)
 DT 01-FEB-1991 (Rel. 18, Last annotation update)
 DE MESSENGYME-SPECIFIC CELL SURFACE GLYCOPROTEIN PRECURSOR (MSP130).
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90110195; PubMed-2295637;
 RA Part B.A., Parks A.L., Raft R.A.;
 RT "Promoter structure and protein sequence of msp130, a lipid-anchored
 sea urchin glycoprotein."
 RL J. Biol. Chem. 265:1408-1413(1990).
 RN (2)
 RP SEQUENCE OF 542-779 FROM N.A.
 RX MEDLINE-87191419; PubMed-3569664;
 RA Leaf D.S., Anstrom J.A., Chin J.E., Harkey M.A., Showman R.M.,
 RA Raft R.A.;
 RT "Antibodies to a fusion protein identify a cDNA clone encoding
 msp130, a primary mesenchyme-specific cell surface protein of the sea
 urchin embryo."
 RL Dev. Biol. 121:29-40(1987).
 CC -1- FUNCTION: NOT KNOWN. COULD BE INVOLVED IN MESENCHYME CELL
 MIGRATION, ADHESION, FUSION, OR SPICULE FORMATION.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (PROBABLY).
 CC -1- TISSUE SPECIFICITY: RESTRICTED TO THE PRIMARY MESENCHYME CELL
 LINEAGE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M31750; AAA30065.1; -
 DR EMBL: M31751; AAA30066.1; -
 DR EMBL: M16457; AAA30064.1; -
 DR PIR: A35006; A35006.
 KW Glycoprotein; Signal; Repeat; GPI-anchor.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 779 MESENCHYME-SPECIFIC CELL SURFACE
 FT GLYCOPROTEIN.
 FT DOMAIN 39 70 GLY-RICH.
 FT FT 262 350 GLY-RICH.
 FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 779 AA; 81006 MW; 7D522D8348928316 CRC64;

Query Match 81.5%; Score 66; DB 1; Length 779;
 Best Local Similarity 78.6%; Pred. No. 2;
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGG 14
 ||||| |||||
 DB 57 GGGGAGRGGGGGG 70

RESULT 3

HOLD3_HOLD1 STANDARD; PRT: 104 AA.

ID HOLD3_HOLD1 STANDARD; PRT: 104 AA.
 AC Q25055;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HOLOTRICIN 3 PRECURSOR.
 OS Holotrichia diomphalia.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Scarabaeiformia; Scarabaeidae; Melolonthinae; Holotrichia.
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-40.
 RC TISSUE-LARVAL, HEMOLYMPH;
 RX MEDLINE-96073722; PubMed-8535393;
 RA Lee S.Y., Moon H.J., Kurata S., Natori S., Lee B.L.;
 RT "Purification and cDNA cloning of an antifungal protein from the
 hemolymph of Holotrichia diomphalia larvae."
 RL Biol. Pharm. Bull. 18:1049-1052(1995).
 CC -1- FUNCTION: HOLOTRICIN 3 HAS ANTIFUNGAL ACTIVITY AGAINST
 C. ALBICANS.
 CC -----
 CC -1- SIMILARITY: TO TENECIN 3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D13744; BAA02889.1; -
 DR HSSP: P30129; ADPV.
 KW Insect immunity; Antibiotic; Hemolymph; Fungicide; Signal; Repeat.
 FT SIGNAL 1 20
 FT PEPTIDE 21 104
 FT DOMAIN 27 98
 FT REPEAT 27 30
 FT REPEAT 31 34
 FT REPEAT 35 38
 FT REPEAT 39 42
 FT REPEAT 43 46
 FT REPEAT 47 50
 FT REPEAT 51 54
 FT REPEAT 55 58
 FT REPEAT 59 62
 FT REPEAT 63 66
 FT REPEAT 67 70
 FT REPEAT 71 74
 FT REPEAT 75 78
 FT REPEAT 79 82
 FT REPEAT 83 86
 FT REPEAT 87 90
 FT REPEAT 91 94
 FT REPEAT 96 98
 SO SEQUENCE 104 AA; 9026 MW; 2799D681BFDCC725 CRC64;

Query Match 80.2%; Score 65; DB 1; Length 104;
 Best Local Similarity 80.0%; Pred. No. 0.46;
 Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGGS 15
 ||||| |||||
 DB 64 GGGGSGGSGGGGGS 78

```

RESULT 4
K2C1_HUMAN STANDARD; PRT; 643 AA.
ID K2C1_HUMAN
AC P04254;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE KERATIN, TYPE II CYTOSKELETAL 1 (CYTOKERATIN 1) (K1) (CK 1) (67 KDA
DE CYTOKERATIN) (HAIR ALPHA PROTEIN).
GN KRT1 OR KRT1A.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-85166239; PubMed-2580302;
RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;
RT "Structure of a gene for the human epidermal 67-kDa keratin.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-85166239; PubMed-2580302;
RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;
RT "Structure of a gene for the human epidermal 67-kDa keratin.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).
RN [3]
RP PRELIMINARY SEQUENCE OF 151-643 FROM N.A.
RX MEDLINE-85207740; PubMed-2581964;
RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA Roop D.R.;
RT "Amino acid sequences of mouse and human epidermal type II keratins
RT of Mr 67,000 provide a systematic basis for the structural and
RT functional diversity of the end domains of keratin intermediate
RT filament subunits.";
RL J. Biol. Chem. 260:7142-7149(1985).
RN [4]
RP REVISIONS, AND VARIANT EHK PRO-160.
RX MEDLINE-92386601; PubMed-1381288;
RA Chiprev C.C., Korge B.P., Markova N., Bale S.J., Diglovaana J.J.,
RA Compton J.G., Steinert P.M.;
RT "A leucine->proline mutation in the H1 subdomain of keratin 1
RT causes epidermolytic hyperkeratosis.";
RL Cell 70:821-828(1992).
RN [5]
RP VARIANTS EHK GLY-154; SER-187 AND PRO-192.
RX MEDLINE-94117869; PubMed-7507151;
RA Yang J.-M., Chiprev C.C., Diglovaana J.J., Bale S.J., Marekov L.N.,
RA Steinert P.M., Compton J.G.;
RT "Mutations in the H1 and 1A domains in the keratin 1 gene in
RT epidermolytic hyperkeratosis.";
RL J. Invest. Dermatol. 102:17-23(1994).
RN [6]
RP VARIANTS EHK PRO-185 AND SER-187.
RX MEDLINE-94117870; PubMed-7507152;
RA McLean W.H.I., Eady R.A., Doppling-Repenstal P.J., McMillan J.R.,
RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA Morley S.M.;
RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT congenital ichthyosiform erythroderma (BCIE).";
RL J. Invest. Dermatol. 102:24-30(1994).
RN [7]
RP VARIANT EHK GLN-489.
RX MEDLINE-92376531; PubMed-1380725;
RA Rothnagel J.A., Dominy A.M., Dempsey L.D., Longley M.A.,
RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT hyperkeratosis.";
RL Science 257:1128-1130(1992).
RN [8]
RP VARIANT ALLELE 1B.
RX MEDLINE-93107743; PubMed-1281859;
RA Korge B.P., Compton J.G., Steinert P.M., Mischke D.;

```

```

RT "The two size alleles of human keratin 1 are due to a deletion in the
RT glycine-rich carboxyl-terminal V2 subdomain.";
RL J. Invest. Dermatol. 99:697-702(1992).
CC -1- SUBUNIT: HETEROTRIMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC KERATIN I IS GENERALLY ASSOCIATED WITH KERATIN 10.
CC TISSUE SPECIFICITY: THE SOURCE OF THIS PROTEIN IS NEONATAL
CC FORESKIN. THE 67-KDA TYPE II KERATINS ARE EXPRESSED IN TERMINALLY
CC DIFFERENTIATING EPIDERMIS.
CC -1- POLYMERIZATION: THERE ARE TWO SIZE VARIANTS OF KRT1, TERMED 1A AND
CC 1B WITH ALLELIC FREQUENCIES OF 0.61 AND 0.39. 1B LACKS 7 LACKS
CC 7 RESIDUES COMPARED TO 1A.
CC -1- DISEASE: DEFECTS IN KRT10 AND KRT1 ARE THE CAUSE OF EPIDERMOLYTIC
CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER
CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC CORNEUM.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF237621; AAF60327.1; -.
DR EMBL; M98776; AAB47721.1; -.
DR PIR; A02950; KR02.
DR PIR; A22940; A22940.
DR HSSP; P05412; 1FOS.
DR AARHUS/CHENT-2DPAGE; 4606; NEPGE.
DR MIM; 139350; -.
DR MIM; 113800; -.
DR INTERPRO; IPR001664; -.
DR INTERPRO; IPR003054; -.
DR PFM; PFM0038; filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; coiled coil; Heptad repeat pattern; Keratin;
KW Disease mutation; Polymorphism; Phosphorylation.
FT INIT_MET 0
FT DOMAIN 1 178
FT DOMAIN 179 488
FT DOMAIN 489 643
FT DOMAIN 179 214
FT DOMAIN 215 233
FT DOMAIN 234 325
FT DOMAIN 326 349
FT DOMAIN 350 488
FT SITE 432 432
FT SITE 432 432
FT DOMAIN 1 150
FT DOMAIN 501 640
FT MOD_RES 65 65
FT VARIANT 154 154
FT VARIANT 160 160
FT VARIANT 185 185
FT VARIANT 187 187
FT VARIANT 192 192
FT VARIANT 311 311
FT VARIANT 329 329
FT VARIANT 357 357

```

```

FT VARIANT 489 489 E -> O (IN EHK).
FT VARIANT 536 536 /FTid=VAR_003861.
FT VARIANT 536 536 G -> C.
FT VARIANT 632 632 /FTid=VAR_003862.
FT VARIANT 632 632 R -> K.
FT VARIANT 559 565 /FTid=VAR_003863.
FT VARIANT 559 565 MISSING (IN ALLELE 1B).
FT SEQUENCE 643 AA; 65886 MM; DE945DC462257850 CRC64;

Query Match 80.2%; Score 65; DB 1; Length 643;
Best Local Similarity 80.0%; Pred. No. 2.2;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGSGGGRAGGGGS 15
Db 594 GGGSGGGRAGGGGS 608

RESULT 5
TWST_MOUSE STANDARD: PRT: 206 AA.
AC P26687;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TWIST RELATED PROTEIN (M-TWIST).
GN TWIST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91122450; PubMed-1840517;
RA Wolf C., Thisse C., Stoetzel C., Thisse B., Gerlinger P.,
RA Perrin-Schmitt F.;
RT "The M-twist gene of Mus is expressed in subsets of mesodermal cells
RT and is closely related to the Xenopus X-tw1 and the Drosophila twist
RT genes";
RL Dev. Biol. 143:363-373(1991).
RN [2]
RP FUNCTION.
RX MEDLINE-98001585; PubMed-9343420;
RA Hamamori Y., Wu H.Y., Sartorelli V., Kedes L.;
RT "The basic domain of myogenic basic helix-loop-helix (bHLH) proteins
RT is the novel target for direct inhibition by another bHLH protein,
RT twist.";
RL Mol. Cell. Biol. 17:6563-6573(1997).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR, WHICH SEEMS TO BE
CC INVOLVED IN THE NEGATIVE REGULATION OF CELLULAR DIFFERENTIATION AND
CC IN THE DIFFERENTIATION OF SEVERAL LINEAGES INCLUDING MYOGENESIS,
CC OSTEOGENESIS, AND NEUROGENESIS. INHIBITS MYOGENESIS BY
CC SUBESTRATING E PROTEINS, INHIBITING TRANS-ACTIVATION BY MYF2, AND
CC INITIATING DNA-BINDING BY MYOD THROUGH PHYSICAL INTERACTION. THIS
CC INTERACTION PROBABLY INVOLVES THE BASIC DOMAINS OF BOTH PROTEINS.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC bHLH PROTEIN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: SUBSET OF MESODERMAL CELLS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (bHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63649; AAA40514.1; -
CC EMBL; M63650; AAA40515.1; -

```

```

DR HSSP: P10085; 1MDY.
DR TRANSFAC; T01635; -
DR MGD; MGI:98872; TWIST.
DR INTERPRO; IPR001092; -
DR INTERPRO; IPR003015; -
DR PFAM; PF00010; HLH; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR Differentiation; Developmental protein; Nuclear protein; DNA-binding;
KM Transcription regulation.
FT DOMAIN 80 102 GLY-RICH.
FT DNA_BIND 112 124 BASIC DOMAIN.
FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT VARIANT 36 36 A -> R (IN CDNA).
FT VARIANT 91 91 G -> P (IN CDNA).
FT SEQUENCE 206 AA; 21198 MM; 618AD8E9BE87C555 CRC64;

Query Match 77.8%; Score 63; DB 1; Length 206;
Best Local Similarity 73.3%; Pred. No. 1.3;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGSGGGRAGGGGS 15
Db 83 GGGGAGGGGGGGGS 97

RESULT 6
GRP2_NICSY STANDARD: PRT: 214 AA.
AC P27484;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 2 PRECURSOR.
GN GRP-2.
OS Nicotiana sylvestris (Wood tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92003709; PubMed-1912512;
RA Obokata J., Ohme M., Hayashida N.;
RT "Nucleotide sequence of a cDNA clone encoding a putative glycine-rich
RT protein of 19.7 kDa in Nicotiana sylvestris.";
RL Plant Mol. Biol. 17:953-955(1991).
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60007; CAA42622.1; -
CC PIR; S1731; KNT2S.
CC HSSP; P15277; 1MJC.
CC INTERPRO; IPR001878; -
CC INTERPRO; IPR002059; -
CC PFAM; PF00033; CSD; 1.
CC PFAM; PF00098; zf-CCHC; 2.
CC PRINTS; PR00050; COLDSHOCK.
CC PRINTS; PR00939; C2HCZNFINGER.
CC PROSITE; PS00352; COLD_SHOCK; UNKNOWN; 1.
CC Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 214 POTENTIAL.
FT CHAIN ? 214 GLYCINE-RICH CELL WALL STRUCTURAL
FT DOMAIN 82 158 PROTEIN 2.
FT DOMAIN 176 195 GLY-RICH.

```

```

SQ SEQUENCE 214 AA; 19746 MW; E28DB84558F2A0AA CMC64;

Query Match
Best Local Similarity 77.8%; Score 63; DB 1; Length 214;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGSGGGRASGGG 14
    ||||| ||| |||
Db 176 GGGGGGGRFGGGG 189

RESULT 7
YPG1_ZYMO
ID YPG1_ZYMO STANDARD; PRT; 265 AA.
AC P30794;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 28.1 KDA PROTEIN IN PGM 5 REGION.
GN ZML.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
CC Zymomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 31821 / ZM4 / CP4;
RX MEDLINE-93308069; Pubmed-8320209;
RA Yomano L.P., Scopes R.K., Ingram L.O.;
RT "Cloning, sequencing, and expression of the zymomonas mobilis
phosphoglycerate mutase gene (pgm) in Escherichia coli.";
RL J. Bacteriol. 175:3926-3933(1993).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09649; AAA71930.1; -
FT DOMAIN 139 185 GLY-RICH.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28086 MW; D7E752E204986D3 CRC64;

Query Match
Best Local Similarity 77.8%; Score 63; DB 1; Length 265;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGSGGGRASGGG 14
    ||||| ||| |||
Db 170 GGGGGGGRFGGGG 183

RESULT 8
HXB3_HUMAN
ID HXB3_HUMAN STANDARD; PRT; 431 AA.
AC P14651; P17484; O95615;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMEOBOX PROTEIN HOX-B3 (HOX-2.7)..
GN HOXB3 OR HOX2G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX Acampora D., D'Esposito M., Faiella A., Pannese M., Miglino E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simone A., Boncinelli E.;

```

```

RT "The human HOX gene family."?
RT Nucleic Acids Res. 17:10385-10402(1989).
RN
RP SEQUENCE FROM N.A.
RA Sauvageau G., Thorleifsdottir U., Hough M.R., Hugo P., Lawrence H.J.,
RA Larmann C., Humphries R.K.;
RT "Regulated expression of Hoxb3 in hematopoietic cells causes
RT defective development of alpha beta T lymphocytes and progressive
RT myeloid proliferation."?
RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE OF 188-253 FROM N.A.
RC TISSUE=PLACENTA:
RX MEDLINE=89378558; Pubmed=2570724;
RA Giampolo A., Acampora D., Zappavigna V., Pannese M.,
RA D'Esposito M., Care A., Falella A., Stornaiuolo A., Russo G.,
RA Simeone A., Boncinelli E., Peschle C.;
RT "Differential expression of human HOX-2 genes along the anterior-
RT posterior axis in embryonic central nervous system."?
RT Differentiation 40:191-197(1989).
RN
RP SEQUENCE OF 188-253 FROM N.A.
RX MEDLINE=90215256; Pubmed=2576652;
RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,
RA Gaudino G., Stornaiuolo A., Caffero M., Falella A., Simeone A.;
RL "Organization of human class I homeobox genes."?
RL Genome 31:745-756(1989).
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT
CC 5-9 WEEKS FROM CONCEPTION.
CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X16667; CA34657.1; -
CC EMBL; U59298; AAD10852.1; -
CC EMBL; X16175; CA34297.1; -
CC PIR: S07543; WJHD2G.
CC PIR: D37042; D37042.
CC HSP: P02833; ISAN.
CC TRANSFAC; T01723; -
CC MIM; 142966; -
CC INTERPRO: IPR001356; -
CC INTERPRO: IPR001827; -
CC PRAM; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEOBOX.
CC PRINTS; PR00025; ANTENNAPEDIA.
CC PROSITE; PS00027; HOMEOBOX_1; 1.
CC PROSITE; PS00032; ANTENNAPEDIA; 1.
CC PROSITE; PS00071; HOMEOBOX_2; 1.
CC DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC Transcription regulation.
KW DOMAIN 129 134 ANTP-TYPE HEXAPEPTIDE.
FT DOMAIN 154 178 GLY-RICH.
FT BIND 188 247 HOMEOBOX.
FT CONFLICT 199 200 OL->HY (IN REF. 2).
SO SEQUENCE 431 AA; 44344 MW; 941706EDCC2975E5 CRC64;

```

```

DB      161 GGGGGGGGGGGG 174

RESULT 9
HXB3_MOUSE
ID      HXB3_MOUSE      STANDARD:      PRT:      433 AA.
AC      P09026; P10285; 061680;
DT      01-NOV-1988 (Rel. 09, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      HOMEOBOX PROTEIN HOX-B3 (HOX-2.7) (MH-23).
GN      HOXB3 OR HOXB-3 OR HOX-2.7.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92258392; PubMed=1582411;
RA      Sham M.H., Hunt P., Nonchev S., Papalopulu N., Graham A.,
RA      Boncinelli E., Krumlauf R.;
RT      "Analysis of the murine Hox-2.7 gene: conserved alternative
RT      transcripts with differential distributions in the nervous system and
RT      the potential for shared regulatory regions.";
RL      EMBO J. 11:1825-1836(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95196953; PubMed=7890121;
RA      Brown W.M., Taylor G.R.;
RT      "The 5'-sequence of the murine Hox-b3 (Hox-2.7) gene and its intron
RT      contain multiple transcription-regulatory elements.";
RL      Int. J. Biochem. 26:1403-1409(1994).
RN      [3]
RP      SEQUENCE OF 152-361 FROM N.A.
RX      MEDLINE=88054465; PubMed=2890503;
RA      Lonai P., Arman E., Czosnek H., Ruddle F.H., Blatt C.;
RT      "New murine homeoboxes: structure, chromosomal assignment, and
RT      differential expression in adult erythropoiesis.";
RL      DNA 6:409-418(1987).
RN      [4]
RP      SEQUENCE OF 181-265 FROM N.A.
RX      MEDLINE=89091992; PubMed=2463210;
RA      Graham A., Papalopulu N., Lorimer J., Mcvey J.H., Tuddenham E.G.D.,
RA      Krumlauf R.;
RT      "Characterization of a murine homeo box gene, Hox-2.6, related to the
RT      Drosophila Deformed gene.";
RL      Genes Dev. 2:1424-1438(1988).
CC      -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC      A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC      SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC      -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on way
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X66177; CAA46951.1; -
DR      EMBL; 002278; AAB60496.1; -
DR      EMBL; M18168; AAA37840.1; -
DR      PIR; S20963; S20963.
DR      PIR; C29585; C29585.
DR      HSSP; P02853; ISAN.
DR      TRANSFAC; T01724; -
DR      MGD; MGI:96184; HOXB3.
DR      INTERPRO; IPR001356; -
DR      INTERPRO; IPR001827; -
DR      Pfam; PF00046; homeobox.1.
DR      PRINTS; PR00024; HOMEOBOX.

```

```

DR      PRINTS; PR00025; ANTENNAPEDIA.
DR      PROSITE; PS00027; HOMEOBOX_1; 1.
DR      PROSITE; PS00032; ANTENNAPEDIA; 1.
DR      PROSITE; PS00071; HOMEOBOX_2; 1.
KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW      Transcription regulation.
FT      DOMAIN 129 134
FT      DOMAIN 154 181
FT      DNA_BIND 191 250
FT      CONFLICT 113 113
FT      CONFLICT 119 119
FT      CONFLICT 152 169
FT      CONFLICT 182 182
FT      CONFLICT 216 217
FT      CONFLICT 330 330
FT      CONFLICT 342 361
FT      SEQUENCE 433 AA; 44353 MW; 9ADB3C922663612A6 CRC64;
Oy      1 GGGGGGGGGGGG 14
Db      162 GGGGGGGGGGGG 175

Query Match      77.8%; Score 63; DB 1; Length 433;
Best Local Similarity 78.6%; Pred No. 2.5;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 10
ID      KDGE_DROME      STANDARD:      PRT:      1454 AA.
AC      009103;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      EYE-SPECIFIC DIACYLGLYCEROL KINASE (EC 2.7.1.107) (RETINAL
DE      DEGENERATION A PROTEIN) (DIGLYCERIDE KINASE) (DGK) (DAG KINASE)..
GN      RDGA OR DGK2.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-CANTON-S; TISSUE-HEAD;
RX      MEDLINE=94068563; PubMed=8248222;
RA      Masai I., Okazaki A., Hosoya T., Hotta Y.;
RT      "Drosophila retinal degeneration A gene encodes an eye-specific
RT      diacylglycerol kinase with cysteine-rich zinc-finger motifs and
RT      ankrylin repeats.";
RL      Proc. Natl. Acad. Sci. U.S.A. 90:11157-11161(1993).
CC      -1- FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE PHOTORECEPTOR. ITS
CC      ABSENCE LEADS TO RHABDOMERE DEGENERATION DUE TO DEFECTIVE
CC      PHOSPHOLIPID TURNOVER.
CC      -1- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +
CC      1,2-DIACYLGLYCEROL 3-PHOSPHATE.
CC      -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC      -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT EYE.
CC      -1- DISEASE: MUTANTS OF THIS GENE ARE CHARACTERIZED BY HAVING
CC      PHOTORECEPTOR CELLS THAT DEVELOP NORMALLY BUT DEGENERATE RAPIDLY
CC      AFTER ECLOSION.
CC      -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC      FAMILY.
CC      -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC      BINDING DOMAINS.
CC      -1- SIMILARITY: CONTAINS 4 ANK REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL, D17315; BAA04135.1; -
 DR FLYBASE; FBgn0003217; rda.
 DR INTERPRO; IPR000756; -
 DR INTERPRO; IPR001206; -
 DR INTERPRO; IPR002110; -
 DR INTERPRO; IPR002219; -
 DR PFAM; PF00609; DAGK; 1.
 DR PFAM; PF00781; DAGK; 1.
 DR PFAM; PF00130; DAG-PE-bind; 1.
 DR PFAM; PF00023; ank; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS50479; DAG-PE_BIND_DOM_1; FALSE_NEG.
 DR PROSITE; PS50081; DAG-PE_BIND_DOM_2; FALSE_NEG.
 DR TRANSFERASE; Kinase; ANK repeat; Repeat; Vision;
 KW Phorbol-ester binding.
 FT DOMAIN 2 6
 FT DOMAIN 24 39
 FT DOMAIN 110 115
 FT DOMAIN 227 231
 FT DOMAIN 758 775
 FT DOMAIN 592 642
 FT DOMAIN 662 719
 FT DOMAIN 807 935
 FT DOMAIN 961 1115
 FT DOMAIN 1317 1451
 FT DOMAIN 1317 1349
 FT REPEAT 1317 1349
 FT REPEAT 1350 1382
 FT REPEAT 1386 1418
 FT REPEAT 1419 1451
 FT MUTAGEN 869 869
 FT SEQUENCE 1454 AA; 159675 MW; CE4C81099FEA16AA CRC64;

Query Match 77.8%; Score 63; DB 1; Length 1454;
 Best Local Similarity 78.6%; Pred. No. 6.9;
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGSGGSRASGGG 14
 DB 760 GCGGAGGAGGAGGGG 773

RESULT 11
 GRP_MAIZE STANDARD; PRT; 157 AA.
 ID GRP_MAIZE P10979;
 AC 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE GLYCINE-RICH RNA-BINDING, ABSICISIC ACID-INDUCIBLE PROTEIN.
 GN RAB15.
 OS Zea mays (Maize).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDOSPERM;
 RX MEDLINE-88288401; PubMed-2969461;
 RA Gomez J., Sanchez-Martinez D., Stiefel V., Rigau J., Puigdomenech P.,
 RA Pages M.;
 RA "A gene induced by the plant hormone abscisic acid in response to
 RT water stress encodes a glycine-rich protein.";
 RL Nature 334:262-264 (1988).
 RN 12
 RP SIMILARITY TO RNA-BINDING PROTEINS.
 RX MEDLINE-89097302; PubMed-2521378;

RA Mortenson E., Dreyfuss G.;
 RA "RNP in maize protein.";
 RL Nature 337:312-312 (1989).
 CC -1- FUNCTION: POSSIBLY HAS A ROLE IN RNA TRANSCRIPTION OR
 CC PROCESSING DURING STRESS.
 CC -1- INDUCTION: BY THE PLANT HORMONE ABSICISIC ACID IN RESPONSE TO
 CC WATER STRESS.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; X12564; CAA31077.1; -
 DR PIR; S04536; S04536.
 DR HSSP; P09651; 1UP1.
 DR MAIZEDB; 69261; -
 DR INTERPRO; IPR000504; -
 DR PFAM; PF00076; rrm; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KW RNA-binding.
 FT DOMAIN 10 15
 FT DOMAIN 49 56
 FT DOMAIN 88 154
 FT SEQUENCE 157 AA; 15438 MW; 8A10592248B60D16 CRC64;

Query Match 76.5%; Score 62; DB 1; Length 157;
 Best Local Similarity 78.6%; Pred. No. 1.3;
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGGSGGSRASGGG 14
 DB 116 GGGGSGGSRREGGG 129

RESULT 12
 GRP_DAUCA STANDARD; PRT; 157 AA.
 ID GRP_DAUCA Q03878;
 AC 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE GLYCINE-RICH RNA-BINDING PROTEIN.
 OS Daucus carota (Carrot).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
 CC euasterids II; Apiales; Apiaceae; Daucus.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. QUEEN ANNE'S LACE;
 RA Sturm A.;
 RA "A wound-inducible glycine-rich protein from Daucus carota with
 RT homology to single-stranded nucleic acid binding proteins.";
 RL Plant Physiol. 99:1689-1692 (1992).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE BIOSYNTHESIS AND PROCESSING OF
 CC HETEROGENEOUS NUCLEAR RNA AND IN THE MATURATION OF SPECIFIC MNAS
 CC IN RESPONSE TO WOUNDING.
 CC -1- INDUCTION: IN RESPONSE TO STRESS BY WOUNDING.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL: X58146; CAA41152.1; -.
DR PIR: S14857; S14857.
DR HSSP: P09651; 10P1.
DR INTERPRO: IPR000504; -.
DR PFAM: PF00076; rrm; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
DR RNA-binding.
KW DOMAIN 8 13 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 47 54 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 86 154 GLY-RICH.
SO SEQUENCE 157 AA; 15718 MW; 73FBD644F51CB633 CRC64;

Query Match
Best Local Similarity 76.5%; Score 62; DB 1; Length 157;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGGG 14
    |||| ||| ||||
Db 103 GGGGSGGRGREGGG 116

RESULT 13
GRP1_ORYSA STANDARD; PRT; 165 AA.
ID P25074;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1 PRECURSOR.
GN GRP-1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. INDICA-IR36;
RX MEDLINE-91370862; PubMed-1716496;
RA "Lel M., Wu R.;
RT "A novel glycine-rich cell wall protein gene in rice.";
RL Plant Mol. Biol. 16:187-198(1991).
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X53596; CAA37665.1; -.
DR PIR: S13385; KNRZG1.
DR HSSP: P30129; 4DPV.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 165 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.
FT DOMAIN 31 159 GLY-RICH.
FT REPEAT 56 62 R2 (TYR-RICH).
FT REPEAT 93 99 R2 (TYR-RICH).
FT REPEAT 132 138 R2 (TYR-RICH).
SO SEQUENCE 165 AA; 13536 MW; E36CE31C3650AC9A CRC64;

Query Match
Best Local Similarity 76.5%; Score 62; DB 1; Length 165;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGGG 15
```

```
Db 111 GGGGSGGGGGGGGG 125
    |||| ||| |||||

RESULT 14
GRP2_SORBI STANDARD; PRT; 168 AA.
ID GRP2_SORBI
AC 099070;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLYCINE-RICH RNA-BINDING PROTEIN 2.
GN GRP2.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. TAMARAN FNK 140; TISSUE-LEAF;
RX MEDLINE-91346715; PubMed-1715211;
RA Cretin C., Puigdomenech P.;
RT "Glycine-rich RNA-binding proteins from Sorghum vulgare.";
RL Plant Mol. Biol. 15:783-785(1990).
CC -1- FUNCTION: POSSIBLY HAS A ROLE IN RNA TRANSCRIPTION OR
CC PROCESSING DURING STRESS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X57662; CAA40862.1; -.
DR PIR: S12312; S12312.
DR HSSP: P09651; 10P1.
DR INTERPRO: IPR000504; -.
DR PFAM: PF00076; rrm; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW RNA-binding.
FT DOMAIN 10 15 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 49 56 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 88 157 GLY-RICH.
SO SEQUENCE 168 AA; 16360 MW; 92F4FD07FE4145D8 CRC64;

Query Match
Best Local Similarity 76.5%; Score 62; DB 1; Length 168;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGGG 14
    |||| ||| ||||
Db 114 GGGGSGGRGREGGG 127

RESULT 15
CHIA_MAIZE STANDARD; PRT; 280 AA.
ID CHIA_MAIZE
AC P29022;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ENDOCHITINASE A PRECURSOR (EC 3.2.1.14) (SEED CHITINASE A).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SEED;
RX MEDLINE-92202208; PubMed-1551872;
```


